1. Introduction
2. Background
   1. Paxos

* An algorithm for implementing fault-tolerant distributed systems.
* The heart is a consensus algorithm – how do we get multiple processes that are each trying to assert/propose a value to agree upon and stick with a single value?
* The safety requirements:
  + Only a single value that has been proposed may be chosen
  + Processes learn about values if and only if they have been chosen
* The protocol does not specify any liveness/convergence requirements.
* There are 3 classes of “agents” that take part in the protocol:
  + Proposers – They propose values to be chosen
  + Acceptors – They choose to or not to accept proposed values
  + Learners – They learn the final, single proposed value that was accepted by the acceptors (not all, just a majority, see below)
  + There are no strict requirements on the mappings between the given processes and these roles
* A proposed value can be considered accepted once a majority of acceptors have accepted it.
* The cornerstone of the algorithm lies in determining how and which value must be accepted.
* From a bird’s eye perspective, the acceptors control the proposers and their proposed values – so the working of the algorithm is driven by acceptors forcing the proposers to propose acceptable values, whilst the design of the algorithm revolves around setting down rules for how to accept values.
* The design considerations for accepting values are as follows (revised as new requirements emerge:
  + 1. An acceptor must accept the first proposal it receives – we must begin somewhere
  + Only a single value must be accepted => we’ll turn this around and instead put the responsibility on the proposers and say – only that value may be proposed repeatedly
  + As a proposer, I can see what values have been accepted while proposing, but I cannot predict what values might be accepted in the future. To this end, I somehow seek to control the future acceptances by extracting promises from acceptors regarding the nature of the same
    - Proposals now have a proposal number. To avoid confusions, different proposals must have different numbers, a global ordering of sort – the implementation left open ended. A suggestion would be to just have proposers choose the numbers from non-overlapping sequences and store the last used number in stable storage.
    - Promise to me, the proposer that you, the acceptor will not accept a proposal with a number lower than mine
    - If you have already accepted a proposal, let me know.
  + Due to this extracted promise, we need to change acceptance rule 1 to: 1a. Acceptors can and must only accept proposals that do not violate promises it has made => accept proposals which have numbers > numbers of proposals to which promises have been made
  + 2. If a proposal with value ‘v’ is chosen, then every higher numbered proposal that is chosen by any acceptor has value ‘v’ – this follows from the requirement that only a single value be chosen in a round of Paxos.
  + This is where the implementation of the algorithm is driven backwards – to ensure that no proposal with a value other than ‘v’ with a proposal number higher than the highest accepted proposal number (with value ‘v’) is accepted, the acceptors force the proposer to only issue proposals with value ‘v’. Hence:   
    2a. If a proposal with value ‘v’ is chosen, then every higher-numbered proposal issued by any proposer has value ‘v’.
  + Now we relax constraint 2a by moving to a majority instead of every acceptor. Hence:   
    2b. For any proposal numbered ‘n’ with value ‘v’ issued, there exists a set ‘S’ consisting of a majority of acceptors such that either:
    - a) no acceptor in S has accepted any proposal numbered less than ‘n’, or
    - b) ‘v’ is the value of the highest numbered proposal among all proposals numbered less than ‘n’ accepted by acceptors in ‘S’
* Putting all this together, the algorithm for a single ‘round’ of Paxos sums up to such:   
  Phase 1.   
  (a) A proposer selects a globally exclusive proposal number ‘n’ and sends a prepare request to a majority of acceptors (it could be all acceptors in the implementation) – this is called a ‘prepare’ request.  
  (b) If an acceptor receives a ‘prepare’ request with number ‘n’ greater than any ‘prepare’ request to which it has already responded, it responds to the request with a promise not to accept any more proposals with number less than ‘n’, and the number ‘n’ and value ‘v’ of the highest number proposal it has accepted (if any).  
  Phase 2.  
  (a) If the proposer receives a response to its prepare request numbered ‘n’ from a majority of acceptors, then it sends an ‘accept’ request to each of those acceptors for a proposal numbered ‘n’ with either the value of the highest numbered proposal it received from the acceptors in response to its prepare request, or if no such value exists, then any value of its choosing.  
  (b) If an acceptor receives an accept request for a proposal numbered ‘n’ >= highest prepare request number it has responded to, then it accepts the proposal.
* A few things to note:
  + a) There is no direct correlation between Phases 1 and 2 in terms of a Phase 1 being sufficient for Phase 2. That is, a proposer ‘P1’ could elicit a response to its prepare request but it might end up racing with another proposer ‘P2’ in that acceptors could end up rejecting P1’s accept requests after accepting its prepare requests because P2 is racing P1 and keeps issuing prepare requests with numbers succeeding P1’s prepare requests.
  + b) A decision is implicitly reached when a majority of the acceptors accept the same value ‘v’ – because using induction and the property of there being at least one common acceptor in the intersection of 2 majorities of acceptors, we can show that the acceptors will force any future proposers into re-proposing the same accepted value.
  + c) There is no limit on the number of proposals that can be made – proposers can abandon proposals mid-flight and reissue proposals of higher numbers if they want.
  + d) There is no guarantee of convergence – the protocol is correct, but may never converge.
* To learn a chosen value, the learners must find out that a majority set of acceptors have accepted a single value. There are multiple ways to do this, the most straightforward of which would be to have every acceptor ack acceptances it makes to every learner.
* Optimizations:
  + The first obvious optimization would be some step to alleviate the non-convergence problem. We could have a “distinguished proposer”, a leader who would be the only one trying to issue proposals, circumventing the race problem.
  + Similarly, we could have a distinguished learner, or a set of them to reduce the number of acks that the acceptors would have to send out once they accept a value.

1. Design and Implementation
2. Experimental Methodology and Results
3. Conclusion